The EMBO Journal vol.6 no.9 pp.2573–2580, 1987

# Multiple sites of alternative splicing of the rat fibronectin gene transcript

**Jean E.Schwarzbauer[1], Ramila S.Patel, Deborah Fonda and Richard O.Hynes**

Center for Cancer Research, Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

[1]Present address: Department of Biology, Princeton University, Princeton, NJ 08544, USA

Communicated by F.E.Baralle

**We describe analyses of the structure and expression of the rat fibronectin gene with particular attention to the 40-kb stretch from the center of the gene which encodes 17 type-III repeating units. Each repeat is precisely separated from its neighbors by introns and most are encoded by pairs of exons. Three repeats are encoded precisely by single exons and two of these (EIIIA and EIIIB) are alternatively spliced in a cell type-specific fashion. A third site of alternative splicing (EIIIB) reported here is similar in expression to the previously described EIIIA segment. Both are excluded from mRNA in liver cells and are, therefore, absent from plasma fibronectin. These two alternative splices, plus a third one (V) reported previously, can occur in all possible combinations giving 12 fibronectin mRNAs from a single gene. These splicing variations account for most but not all of the known fibronectin subunit variants. We report investigations designed to detect other regions of alternative splicing. We also show that the pattern of alternative splicing is somewhat altered on oncogenic transformation.**

*Key words:* fibronectin/alternative splicing/gene structure

## Introduction

Fibronectins are glycoproteins made up of a series of repeating units of three types (Kornblihtt *et al.*, 1985; Skorstengaard *et al.*, 1986) and existing evidence suggests that the fibronectin gene consists of multiple exons which correspond with these repeating structural units (Hirano *et al.*, 1983; Hynes, 1985). In the preceding paper, we report that the type I and II repeats are each encoded by separate exons (Patel *et al.*, 1987). In this paper we report the analysis of the 40 kb of genomic DNA which encode the type-III repeats.

Several questions are at issue here. Earlier work has suggested that type-III repeats are typically encoded by pairs of exons (Odermatt *et al.*, 1985; Oldberg and Ruoslahti, 1986). It is also known that the single fibronectin gene can give rise to multiple different mRNAs by alternative splicing of certain type-III repeats whose exon pattern differs from this norm (Tamkun *et al.*, 1984; Vibe-Pedersen *et al.*, 1984, 1986; Odermatt *et al.*, 1985). We have shown, using segment-specific antibodies, that these alternatively spliced mRNAs encode distinguishable fibronectin subunits which are characteristic of different cell types and forms of fibronectin (Schwarzbauer *et al.*, 1985; Paul *et al.*, 1986). However, the two previously identified regions of alterantive splicing do not account for all the known subunits (Paul *et al.*, 1986). We report here the existence of a third region of alternative splicing of the fibronectin transcript and show that it too is spliced in a cell type-specific fashion. We also investigate the possibility of other regions of alternative splicing and the effects of oncogenic transformation on the pattern of alternative splicing of fibronectin.

## Results

### *Isolation and analysis of genomic and cDNA clones*

A set of four genomic clones (λrFN2, 3, 4 and 5) covering 40 kb of the rat fibronectin gene (Figure 1) was isolated as described (Tamkun *et al.*, 1984; Patel *et al.*, 1987). The clones were analyzed by restriction mapping and by Southern blotting with cDNA probes. cDNA clones were derived from the genomic clones by passage through retroviral vectors (Schwarzbauer *et al.*, 1987; Patel *et al.*, 1987; see Materials and methods).

The cDNA clones and relevant segments of the genomic clones were subcloned and sequenced. The entire cDNA sequence was determined by combining the sequences of these derived cDNA clones with that previously determined from clones isolated from a rat liver cDNA library (Schwarzbauer *et al.*, 1987; Patel *et al.*, 1987; unpublished data). The 5′ and 3′ parts of this sequence encode type-I and -II repeats and are reported in the preceding paper (Patel *et al.*, 1987). The central portion (~5 kb) encodes a series of type-III repeats and is the subject of the present paper.

The amino acid sequence encoded by this block of cDNA sequence is shown in Figure 2. It consists of 17 homologous type-III repeats each ~90 amino acids long plus a 120-amino acid segment (the V segment), which is not homologous with any other sequence. As discussed further below, 15 of the type-III repeats appear always to be present in mRNA and we number them 1–15. The other two are variably included and we refer to them as extra type-III repeats (EIIIA and EIIIB). The EIIIB repeat has not been detected in previous studies of the primary sequence of fibronectin or of cDNA clones. Its sequence, like that of EIIIA, is entirely typical of type-III repeats.

Analysis of the genomic clones encoding the type-III region allowed us to localize many of the exons encoding these repeats. Those which have been located and completely sequenced are shown in Figure 1. In all cases, the sequences of the exons agreed with those of corresponding cDNA clones except that the EIIIA and EIIIB exons were not found in any cDNA clones isolated from the rat liver library. This is because these exons are always spliced out by hepatocytes (see below).

### *Positions of introns*

Seven type-III repeats each have an intron precisely at their ends (Odermatt *et al.*, 1985). We have now analyzed seven more and find that this rule applies also to them (Figure 2). We also noted previously that the intron at the end of each type-III repeat always falls after the first base of a codon and, as shown in Table I, this is true for all 14 repeats analyzed. There is every reason to suspect that the same will be true for repeats III-4 to III-6.

Figures 1 and 2 also show that most, but not all, type-III repeats are interrupted by introns which are in variable positions, both with respect to the amino acid repeats (Figure 2) and with respect

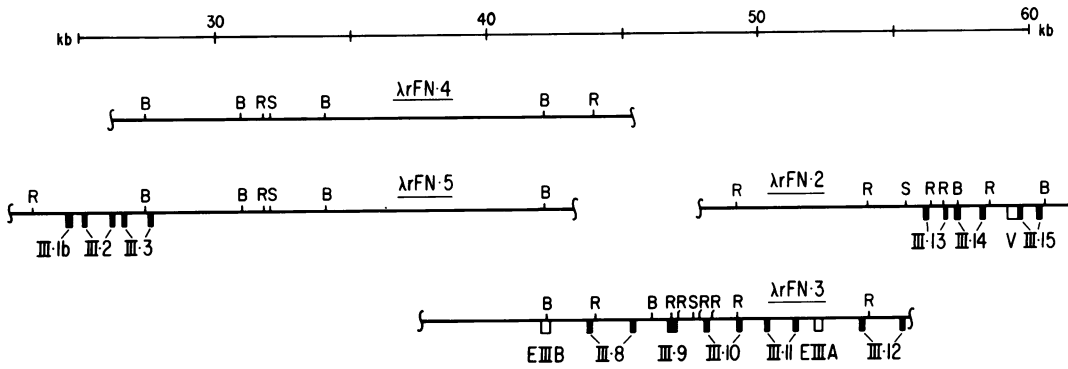


**Fig. 1.** Four rat genomic clones which cover the type-III repeats are shown. The scale is marked in kilobases from the start of the gene (Patel *et al.*, 1987). All *Bam*HI (B), *Eco*RI (R) and *Sal*I (S) sites are marked. Boxes mark exons which have been sequenced. The structure of λ rFN-2 (Odermatt *et al.*, 1985) and the structure of the 5′ and 3′ ends of the gene (Patel *et al.*, 1987) are reported elsewhere.

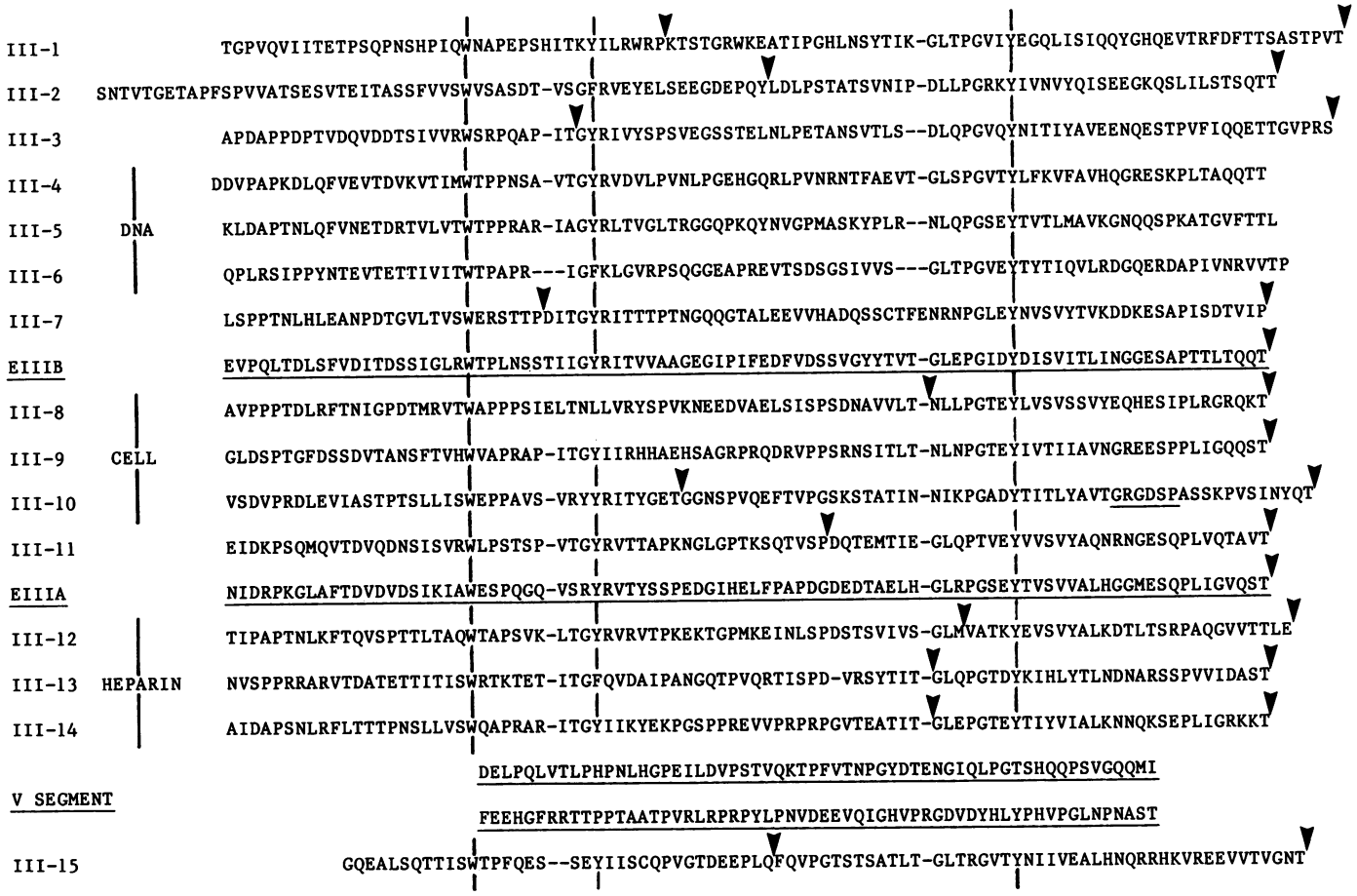


**Fig. 2.** Amino acid sequence of the 17 type-III repeats and the V segment of rat fibronectin. The sequence is given in single letter code and is deduced from the sequence of genomic clones, cDNA clones (Schwarzbauer *et al.*, 1983) and cDNA clones derived from genomic clones (see Materials and methods). The sequence is arranged to display the repeating homologies: conserved aromatic residues are marked by vertical lines. Many other homologies exist but are not marked. Introns are marked by arrowheads (repeats III-4 to III-6 have not been completely analyzed). Note that each repeat is separated from the next by an intron and that most repeats are also interrupted by an intron. Two repeats, EIIIA and EIIIB, are underlined. These repeats are encoded by single exons and are alternatively spliced. Type III-9 is also encoded by a single exon but is not alternatively spliced in any cell type tested. Repeats 10–15 correspond with $n-6$ to $n$ in Odermatt *et al.* (1985).

to codon position (Table II). Three repeats, EIIIA, and EIIIB and III-9, are encoded by larger exons which are not interrupted by introns (Figures 1 and 2, Table II). Thus, most type-III repeats are encoded by a pair of exons of variable size, which together encode a repeat, while three repeats lack the 'middle' intron. In all cases, the repeating structural units are precisely separated from one another in the gene.

### *Alternative splicing and fibronectin variants*

The V region is alternatively spliced in both fibroblasts and hepatocytes (Schwarzbauer *et al.*, 1983, 1985; Paul *et al.*, 1986; Sekiguchi *et al.*, 1986), while the EIIIA segment is always omitted by hepatocytes and is therefore absent from liver mRNA (Kornblihtt *et al.*, 1984a,b, 1985) and from plasma fibronectin

**Table I.** Intron–exon boundaries between type-III repeats

```
                NOT DONE                        CCT GTG ACC A gtacgtatcc
                                          ....III-1....P   V   T

     ttccctccccttttctctag GC AAC ACA            CAG ACT ACA G gtatgtgtgctcc
                          S   N   T....III-2....Q   T   T

   cttcctttctttttctttttag CA CCT GAT            CCA CGA TCC G gtaacttaaaa
                          A   P   D....III-3....P   R   S

                NOT DONE                        GTC ATC CCA G gtaatagaaa
                                          ....III-7....V   I   P

ccccttttgcttcataactcaatag AG GTG CCC           CAG CAA ACG G gtgaatctt
                          E   V   P....EIIIB....Q   Q   T

    ttttccttatttgaaatcag CC GTC CCT            CAG AAA ACA G gtgagccatg.
                          A   V   P....III-8....Q   K   T

cccctccttgggtgggaaaaacag GT CTG GAC           CAA TCC ACG G gtaatgttg...
                          G   L   D....III-9....Q   S   T

    ttttcctgtgatttttctag TT TCC GAT            TAT CAA ACA G gtaaagattt..
                          V   D   D....III-10...Y   Q   T

     ctttcttttgtccttacag AA ATT GAC             GCA GTG ACC A gtacgtaacc..
                          E   I   D....III-11...A   V   T

     ccattaatttgcctaacag AC ATT GAC             CAG TCC ACA G gtatatcggt..
                          N   I   D....EIIIA....Q   S   T

     ccaatgaccatcccgacag CC ATT CTT             ACT CTG GAG A gtgagtaatc..
                          T/A I   P....III-12...T   L   E

     tctgtctgttcctacacag AT GTC AGC             GCC TCC ACG G gtaactaccc..
                          N   V   S....III-13...A   S   T

      ctttgcttgcctttcag CC ATT GAT              AAA AAG ACA G gtaaagactc
                          A   I   D....III-14...K   K   T

      aacctctctcttggctag AT GAG CTT
                          D   E   L....V-25

      tggatgttccctccacag TT CAA AAG
                          V   Q   K....V-95

      atccaaatgcctctacag GA CAA GAA             GGC AAC ACT G gtaggtaacc..
                          G   Q   E....III-15...G   N   T
```

Coding regions are given in upper case letters grouped in codons. Introns are in lower case letters. Amino acids are given in single letter code. Note that all introns between repeats fall at the same position in codons. The first amino acid of repeat III-12 is either threonine (T) or alanine (A) depending on whether EIIIA or III-11 precedes.

**Table II.** Introns interrupting type-III repeats

```
III-1       AGA CCT   gt........ctttgctttatattttgag AAA ACC
             R   P                                  K   T

III-2       TAC CTT G gtgagt..cctttcccctgtatttcaaag  AT CTT
             Y   L                                  D   L

III-3       ATC ACA G gtcatc...cttttctcctaaccctgcag  GG TAC
             I   T                                  G   Y

EIIIB                 INTRON ABSENT

III-8       CTA ACA A gtacgacg...tcccacttgcctcttcag  AT CTC
             L   A                                  N   L

III-9                 INTRON ABSENT

III-10      GAG ACA G gtctgtag......ctccttttacactag  GA GGA
             E   T                                  G   G

III-11      AGT CCA G gtaagaataa.....cctttcccttccag  AT CAA
             S   P                                  D   Q

EIII A                INTRON ABSENT

III-12      CTC ATG   gtaagaagtg......cccttgtttccag GTG GCC
             L   M                                  V   A

III-13      ATT ACA G gtcagtgcgc.........gaattctag  GT TAA
             I   T                                  G   L

III-14      ATC ACT G gtactgcatc......tgtttcccatcag  GT CTG
             I   T                                  G   L

III-15      TTA CAG   gtatatatta.....ctttgggtgtgtag TTC CAA
             L   Q                                  F   Q
```

Coding regions are given in upper case letters grouped in codons. Introns are in lower case letters. Amino acids are given in single letter code.

(Paul *et al.*, 1986). Figure 3B shows nuclease protection experiments which demonstrate that the EIIIB segment is included in mRNA of fibroblasts and astrocytes but is absent from the mRNA of liver, just as is the EIIIA segment (Figure 3A). Therefore, neither EIII segment can contribute to the heterogeneity of pFN, which is synthesized by liver and hepatocytes (Tamkun and Hynes, 1983; Paul and Hynes, 1984; Paul *et al.*, 1986). Thus, EIIIB, like EIIIA, is a cell type-specific segment present in cFN but not pFN. Both are encoded by single exons (Figures 1 and 2, Tables I and II) which can apparently be skipped during splicing of the primary transcript.

Type III-9 is also encoded by a single exon. Can that exon also be alternatively spliced? Figure 3C and D shows that III-9 is always included in mRNAs of normal and transformed fibroblasts, astrocytes and liver cells. The III-9$^-$ form is rare or non-existent in the cell types examined here. Therefore, alternative splicing of this segment does not contribute to the known heterogeneity of fibronectins. It remains possible that III-9 is omitted in some cell type(s) not yet examined.

### *Is there a fourth alternative splice?*

Faced with unexplained fibronectin variants, we considered the possibility that there might be yet another alternatively spliced segment. To explain the known patterns of variants by a single further alternative splice, one would have to invoke a segment which is included approximately half the time by both fibroblasts and hepatocytes (see Discussion). Accordingly, we have screened almost the entire length of fibronectin mRNA in ribonuclease protection experiments. The probes diagrammed in Figure 4 were each tested in hybridization reactions with RNA preparations from various cell types, always including liver and fibroblasts. Representative results are shown in Figure 5. Apart from those which overlap the EIIIB, EIIIA or V segments, none of the probes gave evidence for alternatively spliced variants. The nature of these experiments is such that minor variants can be missed but that a 50:50 mix of two forms (which would protect two different-sized fragments) should generally be readily detectable (see Discussion).

Because repeats III-1 and III-2 are separated by a short non-homologous segment (Kornblihtt *et al.*, 1985; Skorstengaard *et al.*, 1986; see Figure 2), we paid particular attention to the possibility that there might be a special exon, like the V region exon (Tamkun *et al.*, 1984; Vibe-Pedersen *et al.*, 1986), in this region. Nuclease protection experiments covering this region gave no evidence for alternative splicing (Figure 5F and G). We next sequenced the genomic clones encoding this region (Figure 6). The C terminus of III-1 and the N terminus of III-2 each have a non-homologous extension as shown in Figures 2 and 6. However, exons III-1b and III-2a are in two different reading frames. The intron between these exons has terminators in all three reading frames and there is no obvious way in which alternative splicing could produce variations in the segment between the first two type-III repeats (Figure 6). That is, these extensions cannot be alternatively spliced like the non-homologous extension of the V-region exon (Tamkun *et al.*, 1984; Vibe-Pedersen *et al.*, 1986) because they lack appropriate splice sites. The same is true of the non-homologous extensions of repeats III-3 (Figure 6) and III-10 (not shown). Therefore, this detailed

analysis of the rat fibronectin gene does not reveal a fourth region of alternative slicing (see Discussion).

*Alternative splicing in normal and transformed cells*

Previous analyses of protein variants using antibodies against the EIIIA and V segments clearly show that (i) all combinations of EIIIA and V are possible, (ii) different cell types express these two segments in different proportions and (iii) during *in vitro* dedifferentiation of hepatocytes, the splicing of the EIIIA exon changes (Tamkun and Hynes, 1983; Paul and Hynes, 1984; Paul *et al.*, 1986). We next investigated these questions further with particular attention to the alternative splicing of EIIIB. Figure 3 shows that EIIIB, like EIIIA, is included in the mRNA of fibroblasts but not in that of liver. Figure 7 shows that rat fibroblasts contain all four possible combinations of EIIIA and EIIIB, although the EIIIB$^+$ forms ($A^+$ $B^+$, $A^-$ $B^+$) are relatively rare. Early passage rat embryo fibroblasts or astrocytes contain relatively more EIIIB$^+$ mRNA than do established lines (Figure 3B, Figure 7). Figure 7 also shows that the relative proportions of the different forms change on oncogenic transformation. Superimposed on the depression in total fibronectin mRNA levels in transformed cells is a change in the ratio of $A^+$:$A^-$ and $B^+$:$B^-$ forms. The $A^+$ forms represent a slightly higher proportion in transformed cells while the $B^+$ forms are relatively somewhat less common. This was the case whether the transforming virus harbored a p21 Ha-*ras* oncogene (not shown) or a pp60$^{v\text{-}src}$ oncogene (Figure 7) and, in the case of tsRSV-transformed cells, these changes correlated with transformation in temperature-shift experiments (Figure 7).

Therefore, it appears that all combinations of EIIIB, EIIIA and V regions are possible, that different cells splice these segments in different ways and that the patterns of splicing in a given cell type can change in response to physiological stimuli.

## Discussion

Data reported here and elsewhere (Tamkun *et al.*, 1984; Odermatt *et al.*, 1985) confirm that the majority of type-III repeats of rat fibronectin are encoded by pairs of exons while three exceptional repeats are encoded precisely by single exons. In all cases, each type-III is separated from its neighbors by introns; that is, each type-III structural unit is a separate unit in the gene. The introns which interrupt most type-III repeats are in very variable positions (Figure 2). They either arose independently or have moved during evolution, perhaps by a process of exon sliding as suggested for proteases (Craik *et al.*, 1983). Comparison with available data on the positions of introns in the human gene (Vibe-Pederson *et al.*, 1984, 1986; Oldberg and Ruoslahti, 1986) shows that all or most introns are in identical locations in the rat and human fibronectin genes. Oldberg and Ruoslahti (1986) reported an intron in the center of III-8 which does not agree with our data. However, insufficient information was given to be certain that it was a true intron–exon boundary. Vibe-Pedersen *et al.* (1986) report the presence of an ex-

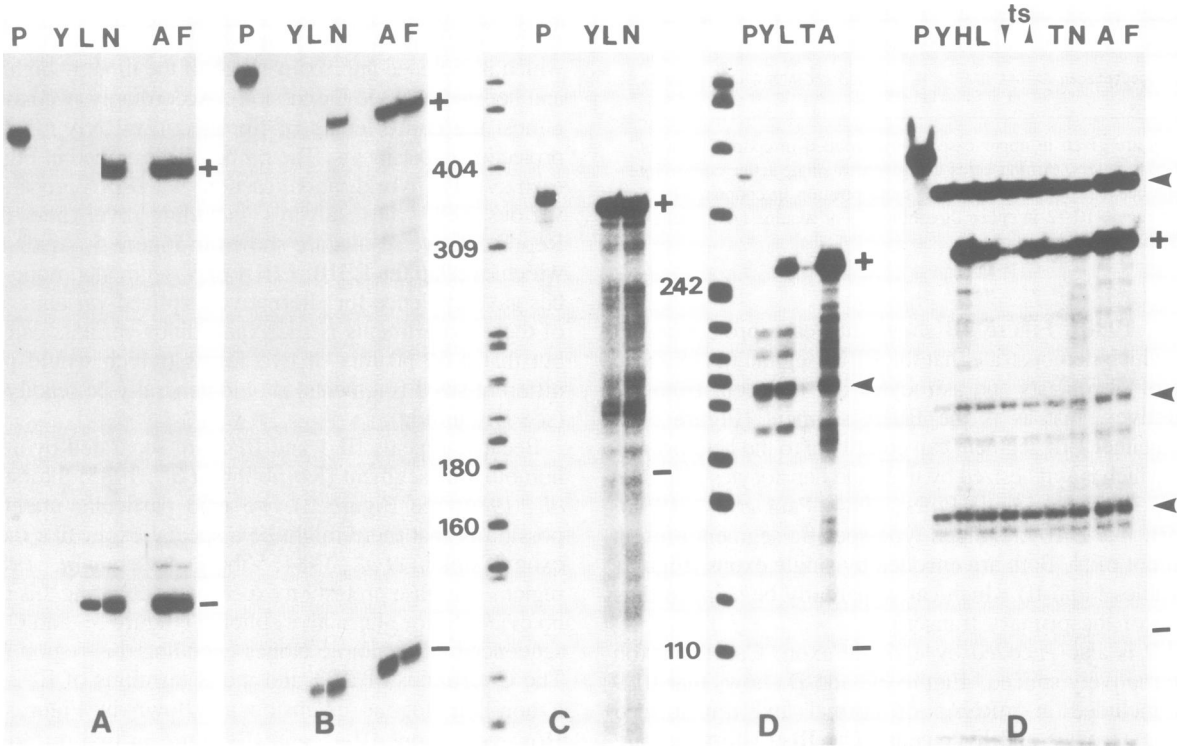


**Fig. 3.** Ribonuclease protection experiments demonstrating alternative splicing of EIIIA and EIIIB but not III-9. In each case the probe (P) designated at the bottom of the panel (see Figure 4) was incubated with RNAs from various sources (Y = yeast control; L = liver; H = hepatocyte; N = rat 1 normal fibroblast cell line; T = RSV-transformed rat 1; A = astrocyte; F = fibroblast; ts = tsRSV-transformed rat 1 at high and low temperatures as indicated by arrowheads). After ribonuclease digestion, protected fragments were analysed on sequencing gels. (**A** and **B**) Probes for EIIIA and EIIIB respectively show protected fragments corresponding to inclusion (+) or omission (−) of these segments in most cell types. Liver RNA contains only EIIIA$^-$ and EIIIB$^-$ forms. (**C** and **D**) Probes crossing the 5′ and 3′ boundaries of repeat III-9 reveal only fragments corresponding to the inclusion of this repeat (+). The predicted sizes of fragments resulting from omission of this repeat are marked by (−). In the case of probe D, there are nuclease-resistant fragments which are independent of the RNA used; see the yeast control lanes (Y). These bands are marked by arrowheads and do not represent specifically protected fragments. The mol. wt markers in **panels C** and **D** are *Msp*I fragments of pBR322 (622, 527, 404, 309, 242, 238, 217, 201, 190, 180, 160, 147, 122 and 110 bp).

tra 5′ splice site (GTGAG) in the exon encoding the V region (called IIIcs in human FN) which is not present in the rat (Tamkun *et al.*, 1984). Apart from these discrepancies, other introns at various points in the human gene correspond precisely with those we have sequenced in the rat gene. Therefore the process of endoduplication and divergence of type-III units occurred long before the divergence of mammals. Consistent with this, the amino acid sequences of individual type-III repeats in rat (Figure 2), human (Kornblihtt *et al.*, 1985) and bovine (Skorstengaard *et al.*, 1986) fibronectins are better than 90% conserved, while different repeats within a species are much less similar (typically 20–40%).

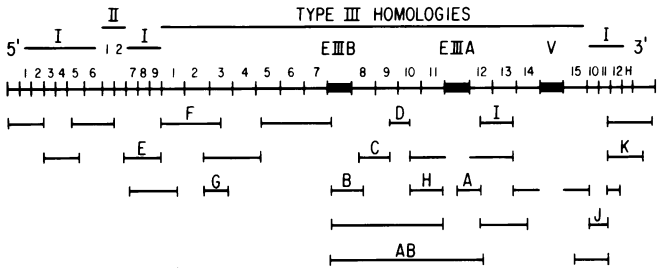


**Fig. 4.** Complementary RNA probes used to screen the fibronectin transcript for alternative splicing. Data for certain of these probes are shown in Figures 3, 5 and 7. Each segment was subcloned in pGEM2 and complementary transcripts prepared with the appropriate polymerase were hybridized with various RNA preparations.

One type-III repeat, EIIIA, is alternatively spliced (Kornblihtt *et al.*, 1984a,b) and is encoded by a single exon (Vibe-Pedersen *et al.*, 1984; Odermatt *et al.*, 1985; Oldberg and Ruoslahti, 1986). We show here that a second type-III repeat, EIIIB, is also encoded by a single exon which is alternatively spliced in a similar fashion. Both EIIIA and EIIIB are always omitted by liver cells, but both can be included by other cell types (Figure 3) and all combinations occur (Figure 7). The EIIIB repeat is rarer than the EIIIA repeat in established cell lines but, in early passage cultures of fibroblasts and astrocytes, the prevalence of EIIIB is similar to that of EIIIA; ~50% in each case. Because of the omission of both EIIIA and EIIIB by hepatocytes, neither repeat occurs in plasma fibronectin (pFN). Inclusion of EIIIA in cellular fibronectin (cFN) produces the larger and more acidic subunits of cFN (Paul *et al.*, 1986). We presume that the same will be true of EIIIB (net negative charge of −6, mol. wt ~10 000 daltons). We have been unable to prepare an antibody reactive with EIIIB and cannot, therefore, test this directly. However, it is reasonable to suggest that $EIIIB^+$ subunits probably comigrate with $EIIIA^+$ subunits of cFN while those subunits which contain both EIIIA and EIIIB probably migrate at an even higher mol. wt and more acidic position. The ribonuclease protection experiments (Figure 7) suggest that the $EIIIA^+/EIIIB^+$ form is rare in established lines but more prominent in early passage cultures. The same is true of the largest and most acidic cFN subunits (J.I.Paul and R.O.Hynes, unpublished results). We have observed EIIIB in chickens (P.Norton and R.O.Hynes, unpublished results) and others have recently detected it in humans

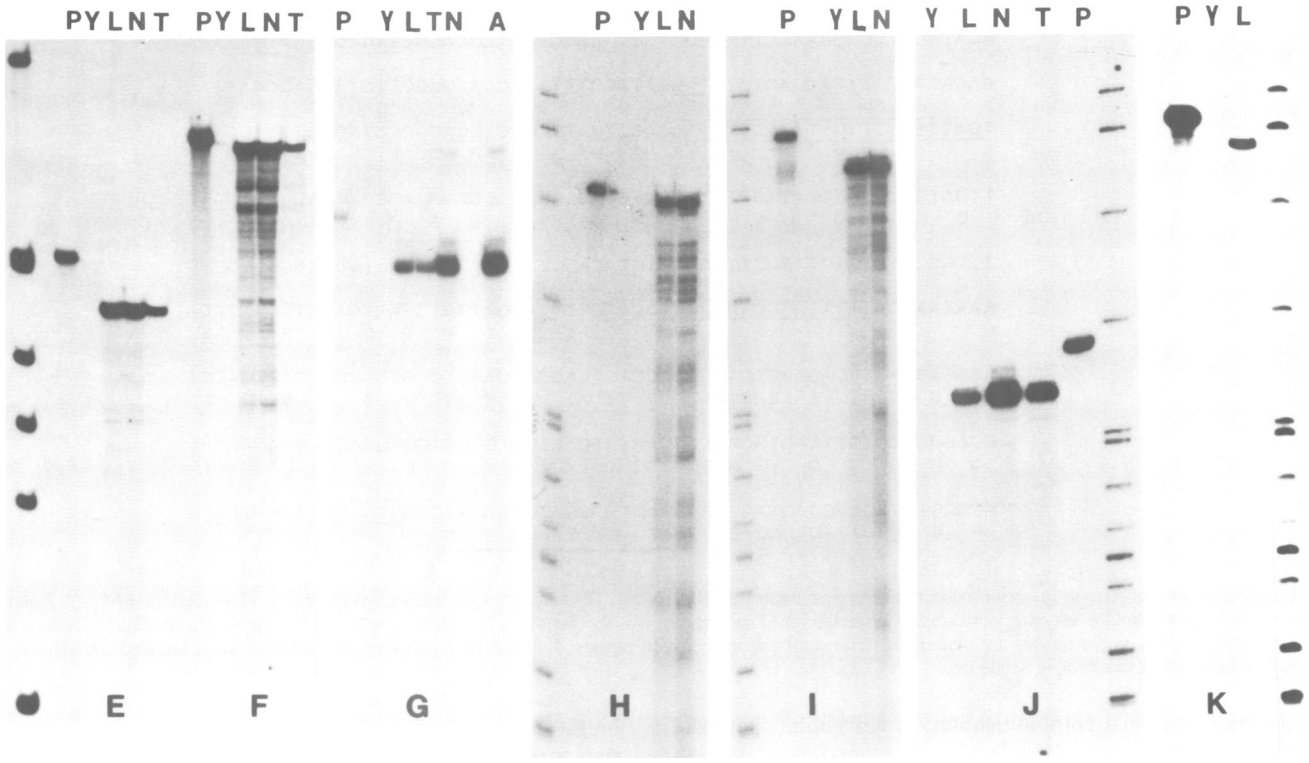


**Fig. 5.** Ribonuclease protection experiments. The probe designated at the bottom of each panel (see Figure 4) was incubated with RNAs from different sources (Y = yeast control; L = liver; N = rat 1 normal fibroblast cell line; T = RSV-transformed rat 1; A = astrocyte). After ribonuclease digestion, protected fragments were analyzed on sequencing gels in parallel with undigested probe (P) and mol. wt markers (pBR322 *Hin*fI fragments in **panels E** and **F**; 1631, 517, 506, 396, 344, 298, 220 bp or pBR322 *Msp*I fragments in other panels; see Figure 3 legend). The repeat boundaries monitored in each experiment can be seen by comparison with Figure 4. In no case were fragments observed which suggested alternative splicing at any of these boundaries. The minor bands in **panels F** and **H** do not correspond with positions which could reflect alternative splicing. They represent background bands which are sometimes seen in these experiments. In particular, in **panel F**, none of these bands are of the sizes predicted for alternative splicing between repeats III-1 and III-2 and III-3.

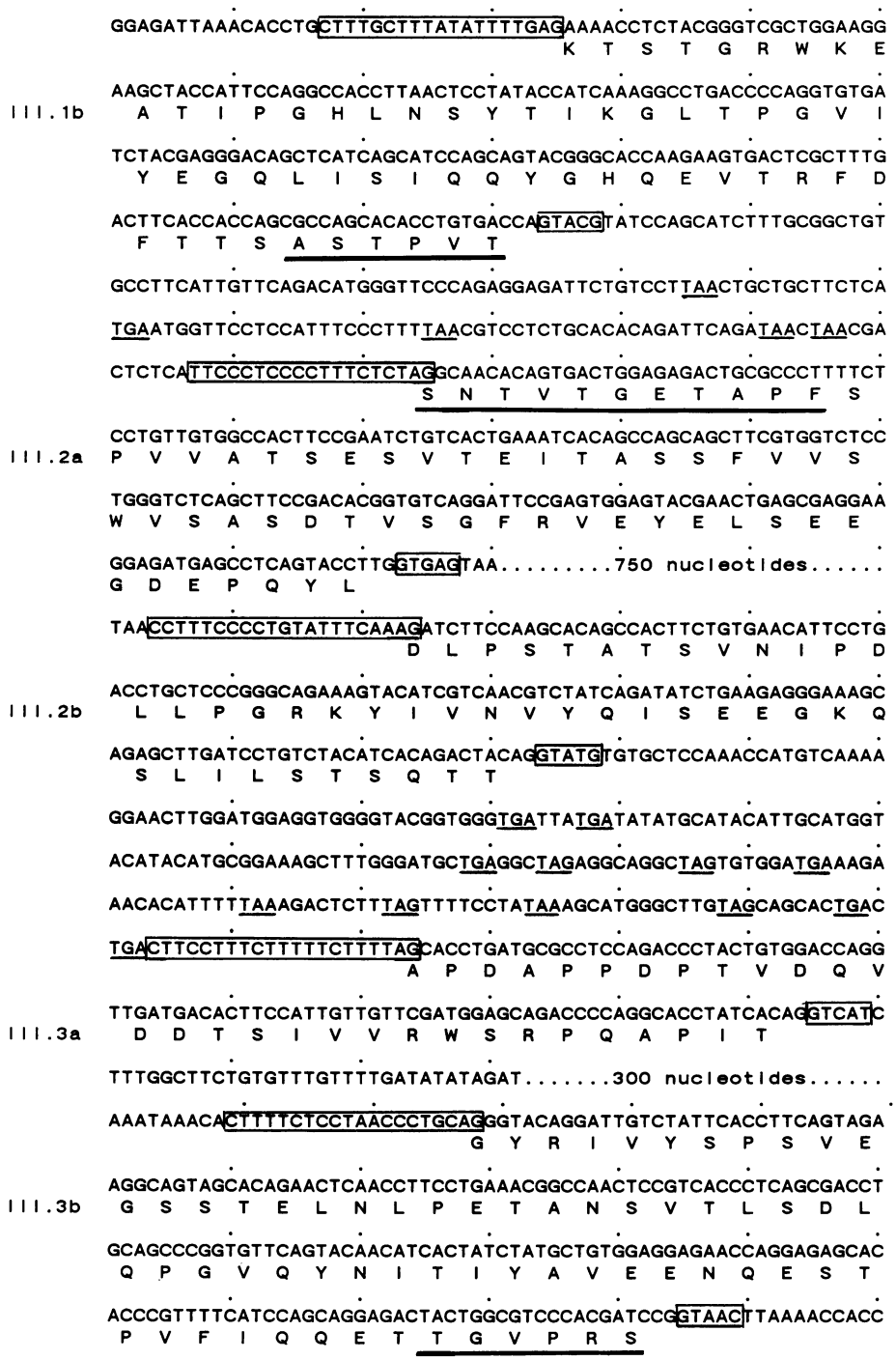


**Fig. 6.** Structure of the gene in the region encoding the first three type-III repeats. The sequences of five exons are shown. The ends of III-1, III-2 and III-3 each have short non-homologous extensions (heavy underlining). The sequence shows that these are unlikely to be alternatively spliced. Splice sites are boxed and termination codons in the introns are underlined. The short introns between repeats III-1 and III-2 and III-3 are shown in their entirety and cannot obviously encode any extra protein sequence.

(F.E.Baralle, personal communication; A.Kornblihtt, personal communication).

The relative proportions of EIIIA$^+$ and EIIIB$^+$ forms of fibronectin mRNA differ between cell types. Furthermore, they change on oncogenic transformation in a reproducible fashion. EIIIB$^+$ forms become even rarer, while the relative proportion of EIIIA$^+$ forms increases somewhat (Figures 3 and 7). The significance of this change is obscure since the functions of these variable segments of fibronectin are unknown. However, it is possible that this change in splicing of fibronectin has a bearing on the loss of fibronectin which is characteristic of oncogenic transformation. Hakomori, Zardi and their colleagues have reported data using monoclonal antibodies which suggest that transformed cells contain greatly elevated levels of certain fibronectin epitopes (Matsuura and Hakomori, 1985; Sekiguchi *et al.*, 1985; Castellani *et al.*, 1986; Borsi *et al.*, 1987). They propose that these epitopes lie in alternatively spliced segments, specifically the EIIIA and V regions. These data are not all in accord with ours since we see a decrease in EIIIB content and only a modest increase in representation of EIIIA. Furthermore,

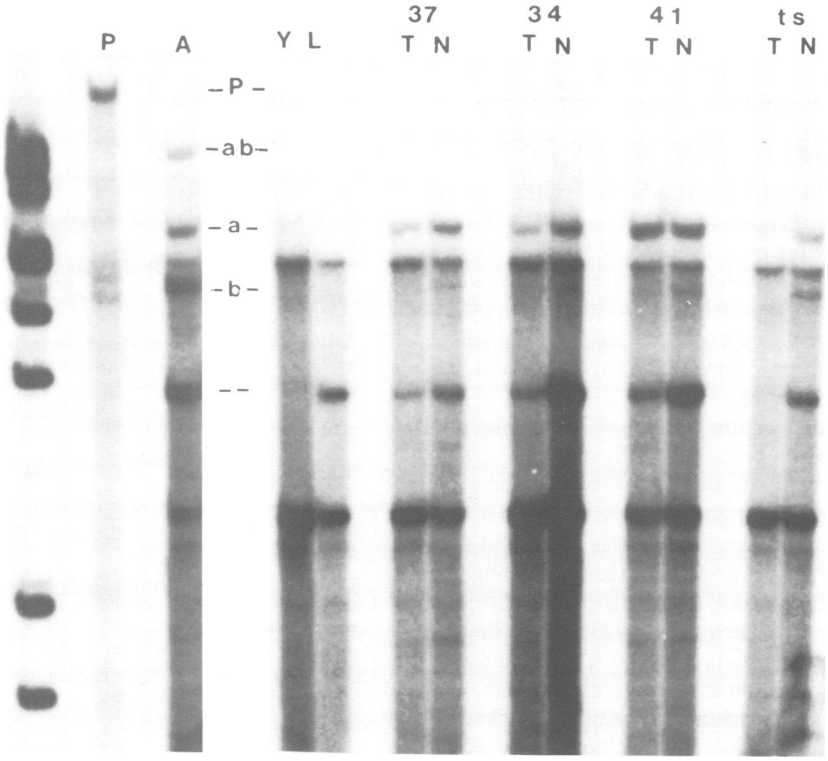


**Fig. 7.** Ribonuclease protection experiments showing all combinations of EIIIA and EIIIB splice variants. The long probe AB (Figure 4) which extends from EIIIB to EIIIA was incubated with RNAs from different sources. After ribonuclease digestion, the protected fragments were analyzed on a 5% gel. Long probes such as this one (P) give background nuclease-resistant bands independent of the RNA used for protection. These are present in the yeast control lanes (Y) as well as the experimental samples and should be ignored. Fragments corresponding to inclusion of both EIIIA and EIIIB (ab), either one alone (a or b) or neither (−) are marked. Note that RNA from early passage culture of astrocytes (A) shows significant amounts of $EIIIA^+/EIIIB^+$ mRNA as well as the other forms. RNAs from cultured cell lines (T and N) also show all four forms, although the $EIIIB^+$ forms (ab,b) are rarer. Liver RNA (L) shows only the $EIIIA^-/EIIIB^-$ form. Normal rat 1 fibroblasts (N) and their RSV-transformed derivatives (T) grown at various temperatures show that the relative frequency of inclusion of EIIIA rises on transformation (compare ratios of bands marked a and −) while the inclusion of EIIIB falls (note bands ab and b).

fibroblastic cells make predominantly $V^+$ forms (Tamkun and Hynes, 1983; Paul and Hynes, 1984; Paul *et al.*, 1986), so there is no chance for a large increase in representation of the V segment on transformation. At present we cannot explain the discrepancy between our nucleic acid results and the implications of the antibody data. One possibility is that the epitopes being recognized are not primary sequence epitopes.

Thus, both the EIIIA and EIIIB exons are alternatively spliced in a cell type-specific fashion which is subject to modulation during physiological changes such as transformation (Figures 3 and 7) and dedifferentiation of hepatocytes.

We also report here that type III-9 is encoded by a single exon. However, we do not detect alternative splicing of this exon in any of the cell types tested so far which include liver, primary hepatocytes, early passage fibroblasts and astrocytes and both normal and transformed fibroblastic cell lines (Figure 3D). It remains possible that this exon is alternatively spliced in some cell type or situation we have not investigated. Therefore, the fact that a type-III repeat is encoded by a single exon is not sufficient for alternative splicing. It is clear that cells can discriminate among the intron−exon boundaries within the fibronectin transcript in a complex fashion. The transcript is probably unique, with single initiation and poly(A) addition sites (Patel *et al.*, 1987), so some local features of the transcript are recognized by the splicing machinery to treat the EIIIA and EIIIB exons differently. Furthermore, different cell types do so in different ways. Casual inspection of the sequences of the intron−exon boundaries (Table I) does not reveal any unique features of the EIIIA and EIIIB boundaries. The only obviously unusual boundary is the 5′ boundary of III-9 which has an unusually purine-rich 3′ splice acceptor sequence.

The three known segments of alternative splicing (EIIIB, EIIIA, V) cannot account fully for the heterogeneity in fibronectin subunits detectable on 2D gels (Paul *et al.*, 1986). Known post-translational modifications (N- and O-linked sugars, sulfation and phosphorylation) also fail to account for the unexplained heterogeneity (Paul and Hynes, 1984; Paul *et al.*, 1986; J.Paul and R.O.Hynes, unpublished data). Therefore, we have investigated in some detail the possibility that a fourth region of alternative splicing may exist. The most obvious possibilities were the type III-9 exon and the non-homologous segment between repeats III-1 and III-2. As discussed above, III-9 is not alternatively spliced in any cell type tested and, as shown in Figures 5 and 6, III-1 and III-2 are not alternatively spliced in any fashion similar to the V region. A systematic screen of almost the entire fibronectin sequence by ribonuclease protection experiments also failed to reveal any other regions of alternative splicing (Figures 4 and 5). The probes used (Figure 4) cover most of the boundaries between repeats except those between I-2/I-3 and II-1/II-2. However, no data suggest the existence of heterogeneity in the fibrin- or gelatin-binding domains which contain these two boundaries. All boundaries between type-III repeats have been screened and, with the exception of boundaries flanking EIIIB, EIIIA and V, we have not detected evidence for alternative splicing. We cannot rule out minor splice variants because of background bands in some of the nuclease protection experiments (Figures

3 and 5). However, no alternative splice forms of the prevalence (~50:50) necessary to explain the other variant fibronectin subunits were detected. One possibility which would be missed in these experiments would be further variation occurring precisely at the boundaries of EIIIB, EIIIA or V. We have sequenced in their entirety the introns preceding V (Tamkun *et al.*, 1984), and following EIIIB (unpublished data) and much of the introns preceding EIIIB (unpublished results) and flanking EIIIA (Odermatt *et al.*, 1985) without uncovering evidence for any other unknown exons. We have not detected any long open reading frames and certainly no other type-III repeats. It remains possible that further alternatively spliced exons are lurking somewhere within the 70 kb of the fibronection gene but, at present, no evidence points to their existence.

In conclusion, the rat fibronection gene is ~70 kb long and contains almost 50 exons. Each of the repeating units of the protein is encoded by a separate exon or pair of exons. The gene is transcribed probably from a single initiation site to a single poly(A) addition site and this 70-kb transcript is spliced to give mRNAs of ~8 kb. At three positions, alternative splicing occurs, giving rise to 12 different mRNAs which encode different fibronectin subunits (Schwarzbauer *et al.*, 1983, 1985; Paul *et al.*, 1985). This alternative splicing is cell type-specific and subject to modulation by transformation and differentiation state. The variations produced account for much but not all of the known heterogeneity of fibronectin subunits. The mechanisms by which the alternative splicing is regulated are unknown and the functional significance of the segments which are variably included is unknown. These questions offer fruitful ground for future research.

## Materials and methods

*Origin of genomic and cDNA clones*

Isolation of genomic clones is described elsewhere (Tamkun *et al.*, 1984; Patel *et al.*, 1987). Clones λ rFN2 to λ rFN5 cover the central part of the gene which includes all the type-III repeats (Figure 1). cDNA clones isolated from a rat liver λ gt11 library cover type-III repeats 9–15 (Schwarzbauer *et al.*, 1983). cDNA clones covering type-III repeats 1–9 and the two alternatively spliced EIII repeats were obtained by passage of genomic clones through retroviral vectors (Schwarzbauer *et al.*, 1987; Patel *et al.*, 1987).

*Analysis of clones*

Relevant fragments were subcloned into pGEM vectors (Promega Biotec) and sequenced by the method of Maxam and Gilbert (1980) or by dideoxy sequencing using oligonucleotide primers, Klenow fragment of DNA polymerase or reverse transcriptase and the conditions suggested by Promega Biotec.

*Nuclease protection analysis*

RNAs were prepared by the guanidine thiocyanate procedure (Chirgwin *et al.*, 1979). Complementary RNA probes were prepared using cDNA subclones in pGEM vectors transcribed with SP6 or T7 polymerase (Promega Biotec) and purified in acrylamide gels. Five micrograms total RNA plus 5 μg yeast tRNA were incubated with $1-3 \times 10^6$ c.p.m. of probe in 30 μl 80% formamide, 40 mM Pipes pH 6.4, 400 mM NaCl, 1 mM EDTA overnight at 37°C. Excess unhybridized probe was digested with 2 μg/ml T1 ribonuclease and 40 μg/ml pancreatic ribonuclease in 10 mM Tris pH 7.5, 300 mM NaCl, 5 mM EDTA for 30 min at 22°C. The nucleases were removed by digestion with proteinase K and extraction with phenol–chloroform. Protected fragments were analyzed on sequencing gels.

## Acknowledgements

This work was supported by a grant for the USPHS National Cancer Institute (POI CA26712) and by grants to the M.I.T. Cancer Center from the National Cancer Institute, Bristol Meyers Inc. and Ajinomoto, Inc. J.E.S. was a fellow of the Charles A. King Trust. We thank Pam Norton for helpful criticism of the manuscript.

## References

Borsi,L., Carnemolla,B., Castellani,P., Rosellini,C., Vecchio,D., Allemanni,G., Chang,S.E., Taylor-Papadimitriou,J., Pande,H. and Zardi,L. (1987) *J. Cell Biol.*, **104**, 595–600.

Castellani,P., Siri,A., Rosellini,C., Infusini,E., Borsi,L. and Zardi,L. (1986) *J. Cell Biol.*, **103**, 1671–1677.

Chirgwin,J.M., Przybyla,A.E., MacDonald,R.J. and Rutter,W.J. (1979) *Biochemistry*, **18**, 5194–5299.

Craik,C.S., Rutter,W.J. and Fletterick,R. (1983) *Science*, **220**, 1125–1129.

Hirano,H., Yamada,Y., Sullivan,M., deCrombugghe,B., Pastan,I. and Yamada,K.M. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 46–50.

Hynes,R.O. (1985) *Annu. Rev. Cell Biol.*, **1**, 67–90.

Kornblihtt,A.R., Vibe-Pedersen,K. and Baralle,F.E. (1984a) *EMBO J.*, **3**, 221–226.

Kornblihtt,A.R., Vibe-Pedersen,K. and Baralle,F.E. (1984b) *Nucleic Acids Res.*, **12**, 5853–5868.

Kornblihtt,A.R., Umezawa,K., Vibe-Pedersen,K. and Baralle,F.E. (1985) *EMBO J.*, **4**, 1755–1759.

Matsuura,H. and Hakomori,S. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 6517–6521.

Maxam,A.M. and Gilbert,W. (1980) *Methods Enzymol.*, **65**, 499–580.

Odermatt,E., Tamkun,J.W. and Hynes,R.O. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 6571–6575.

Oldberg,A. and Ruoslahti,E. (1986) *J. Biol. Chem.*, **261**, 2113–2116.

Patel,R.S., Odermatt,E., Schwarzbauer,J.E. and Hynes,R.O. (1987) *EMBO J.*, **6**, 000–000.

Paul,J. and Hynes,R.O. (1984) *J. Biol. Chem.*, **259**, 13407–13487.

Paul,J.I., Schwarzbauer,J.E., Tamkun,J.W. and Hynes,R.O. (1986) *J. Biol. Chem.*, **261**, 12258–12265.

Schwarzbauer,J.E., Tamkun,J.W., Lemischka,I.R. and Hynes,R.O. (1983) *Cell*, **35**, 421–431.

Schwarzbauer,J.E., Paul,J.I. and Hynes,R.O. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 1424–1428.

Schwarzbauer,J.E., Mulligan,R.C. and Hynes,R.O. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 754–758.

Sekiguchi,K., Siri,A., Zardi,L. and Hakomori,S.I. (1985) *J. Biol. Chem.*, **260**, 5105–5114.

Sekiguchi,K., Klos,A.M., Kurachi,K., Yoshitake,S. and Hakomori,S. (1986) *Biochemistry*, **25**, 4936–4941.

Skorstengaard,K., Jensen,M.S., Sahl,P., Petersen,T.E. and Magnusson,S. (1986) *Eur. J. Biochem.*, **161**, 441–453.

Tamkun,J.W. and Hynes,R.O. (1983) *J. Biol. Chem.*, **258**, 4641–4647.

Tamkun,J.W., Schwarzbauer,J.E. and Hynes,R.O. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 5140–5144.

Vibe-Pedersen,K., Kornblihtt,A.R. and Baralle,F.E. (1984) *EMBO J.*, **3**, 2511–2516.

Vibe-Pedersen,K., Magnusson,S. and Baralle,F.E. (1986) *FEBS Lett.*, **207**, 287–291.

*Received on April 7, 1987; revised on June 5, 1987*

## Note added in proof

The cDNA sequence and parts of the genomic sequence have been submitted to the GenBank and EMBL databases.